

Luci e ombre di Google. Futuro e passato dell'industria dei metadati

Salvatore Poier

<salvatore.poier@unimi.it>

Copyright © 2007 Salvatore Poier

Questo documento è soggetto a una licenza [Creative Commons](#)

24-04-2007

Luci e ombre di Google è uno di quei rari libri che consente la comprensione di un fenomeno esteso sotto molteplici aspetti. L'analisi del complesso di Mountain View, infatti, non è solo politica e ideologica, bensì procede di pari passo con la spiegazione del funzionamento tecnico di Google, per il quale alcune informazioni sono immediatamente reperibili, mentre altre semplicemente ignorate. Il volume si apprezza anche per la pluralità delle voci che lo compongono: non è, infatti, opera di un autore, bensì di un collettivo di appassionati di elettronica, informatica, filosofia ed etica *hacker* che si riconosce e ritrova nel sito www.ippolita.net.

Il gruppo Ippolita, dopo aver presentato la prima bozza in occasione di Hackit, svoltosi a Parma nel 2006, ha dato alla luce un prodotto che non considera finito, ma che si configura come una cristallizzazione di pensieri e di considerazioni esposte a critiche e revisioni, per la natura stessa del *medium* della rete. Più che un libro su Google, *Luci e ombre di Google* è un libro su un metodo - quello *hacker* - di gestione e scambio delle informazioni. Google è un pretesto - come lo fu l'*Open Source* per il precedente *Open non è free*, Elèuthera 2005 (anch'esso rinvenibile gratuitamente sul [sito](#) - per illustrare ancora una volta uno stile di ricerca che pone il soggetto, le sue preferenze ed esigenze, sopra a tutto, e che guarda con sospetto chi, pretendendo di essere *super partes*, spesso si rivela accentratore di nuovi poteri e saperi.

Il volume narra la storia di Internet dal punto di vista dei motori di ricerca. Narra cioè la storia di un luogo senza confini che è modellato e creato costantemente dai percorsi dei singoli, che diventano così soggetti performanti di Internet stessa. Essendo i percorsi potenzialmente infiniti, e infinitamente modificabili, è palese quanto importante sia il ruolo di un motore di ricerca. A causa dell'enorme quantità di dati da analizzare, è fondamentale impostare dei filtri che indicizzino le informazioni nel modo più efficace. Le caratteristiche e le impostazioni dei filtri, il cuore dei motori di ricerca, rendono quindi il ruolo di questi ultimi tutt'altro che neutrale. Il risultato di una ricerca effettuata dall'utente non sarà uno sterile resoconto di occorrenze della parola cercata, bensì risponderà a un complesso e sofisticato algoritmo.

Una serie di scelte fortunate ha portato Google ad essere quello che oggi è: il più usato motore di ricerca al mondo. Il segreto è una sapiente miscela di ingredienti ben dosati: un buon algoritmo di indicizzazione, un *hardware* potente, un'etica improntata accademicamente, un uso sapiente delle comunità *Open Source* e delle risorse che da queste derivano. L'Autore illustra con competenza tutti questi ingredienti, non fermandosi agli aspetti tecnici, bensì rapportandoli costantemente ad un'approfondita analisi storica e politica, dando quindi una panoramica completa di quanto succede attorno ad Internet e delle lotte di potere che si stanno combattendo a suon di protocolli di comunicazione tra computer e di *software*. Google non si limita più, infatti, a cercare per l'utente nella Rete, ma sta gradualmente proponendo il proprio *software*, i propri parametri e il proprio metodo, imponendo di fatto nuovi *standard*.

Il nucleo del meccanismo è l'algoritmo di ricerca. Un algoritmo è un «insieme di regole o direttive atte a fornire una risposta specifica a uno o più dati in input» (p. 41). L'algoritmo usato da Google si chiama *PageRank*; fu frutto della passione e dell'impegno dei due fondatori di Google, Larry Page e Sergey Brin (anche se la paternità è ascrivibile al primo), ed è un'applicazione delle ricerche statistiche su sistemi chiusi compiute da Andrej Andreevic Markov all'inizio del '900. Page e Brin, conosciutisi a Stanford nel 1995, coltivarono assieme il sogno di creare uno strumento in grado di indicizzare tutte le pagine web (*googol* in matematica indica la cifra 1 seguita da 100 zeri, e *goggles*, informalmente, occhiali, specialmente quelli da protezione). Ma i due volevano dal loro algoritmo un risultato più sofisticato, ossia che non offrisse semplicemente una lista di pagine web nelle quali comparisse la parola cercata, o le parole semanticamente affini, bensì una lista di pagine "giuste", ossia con ogni probabilità vicine all'intenzione dell'utente. Per fare ciò occorreva un filtro potente che ordinasse le pagine in base a un sistema di classificazione efficiente, il quale assicurasse - almeno secondo Page e Brin - risposte pertinenti alle domande e quindi anche una maggior probabilità di incontrare le aspettative dell'utente. Il *PageRank* assegna un voto (*to rank*=classificare), un valore alle pagine indicizzate sulla base del numero di volte che queste pagine sono linkate - cioè citate - da altre pagine web. Il metodo è di evidente natura accademica e si basa sul principio del *peer-review* tipico delle pubblicazioni scientifiche. Se una pagina è citata spesso e da molti siti - e i siti più prestigiosi avranno un peso maggiore nell'assegnazione del voto - allora significa che ha un certo rilievo all'interno della comunità e quindi, presumibilmente, corrisponderà all'oggetto che l'utente sta cercando. Quando, quindi, cercherò "artiglio" in Google, la risposta che riceverò sarà non una sterile lista di tutti i siti che contengono la parola "artiglio", bensì una lista dei siti meglio integrati nella Rete, ossia che hanno acquisito una certa autorevolezza, dimostrata dal fatto che sono frequentemente citati da altri. Quello che *PageRank* effettua è un'analisi del *network* Internet, piuttosto che una ricerca testuale. Secondo i fondatori di Google, questo metodo di classificazione è il più democratico possibile poiché mette in risalto i siti maggiormente visitati, cliccati e linkati, trattanto i *link* come voti. A un maggior numero di *link* corrisponde così un maggior numero di voti, e quindi un consenso maggiore. Questa presunzione è ovviamente fallace, ma è efficace nella maggior parte dei casi. Quello che cerchiamo è, spesso, esattamente tra i primi risultati di Google. A supporto di questo risultato intervengono i *cookies* che Google installa sulle macchine degli utenti e che "profilano" così la risposta del motore di ricerca in base alle preferenze man mano esplicitate dall'utente in occasione di altre ricerche. Se, quindi, un dato giorno - o mese, o anno: i *cookies* di Google non hanno scadenza - avrò ricercato "cibo per gatti", molto probabilmente il risultato della mia ricerca sull'artiglio porrà tra i primi siti quelli inerenti ai gatti e alla loro unghia retrattile, piuttosto che siti riguardanti uccelli rapaci.

L'Autore deve, a questo punto del libro, prendere posizione nei confronti di Google. Se, infatti, è vero che i programmi, le applicazioni e i servizi di Google hanno facilitato di molto la vita degli utenti di Internet, tanto da creare veri e propri *fan club* del Gigante Buono (vedi ad esempio [Unofficial Google Fan Club](#)), è anche vero che ci sono scottanti problemi di *privacy*, a causa dei dati sugli utenti custoditi nei suoi dischi rigidi. Ma c'è di più. Secondo l'Autore la buona novella con cui Google si presenta al mondo, coadiuvata da una interfaccia grafica sobria e rassicurante, porta in realtà ad un atteggiamento passivo degli utenti nel reperimento delle informazioni. Nel *mare magnum* che è Internet riusciamo a trovare solo quelle informazioni che vengono reperite da Google e consideriamo esistenti solo quelle pagine che vengono allo scoperto grazie al *PageRank*.

Nelle ricerche della vita quotidiana tale chiusura assoluta è molto utile, perché conduce rapidamente a un risultato; in realtà però ci illude che la libertà consista nell'ottenere una qualità totale. Sappiamo invece che in un sistema reticolare complesso non esistono verità assolute, ma solo autorità distribuite a seconda del percorso che si desidera affrontare, o anche solamente in funzione del tempo che si è disposti a investire nella ricerca. La qualità dipende interamente dalla nostra soggettiva percezione dell'accettabilità del risultato. (p. 50)

Per Ippolita, però, Google non è un male in sé; dobbiamo soltanto essere consapevoli che esso è uno strumento che deve essere usato con cautela. Una ricerca, infatti, è un processo, e come tale può anche guidarci verso lidi completamente inaspettati, “sbagliati” per certi versi, perché non conducono direttamente a quanto stavamo cercando, ma proprio per questo estremamente “giusti”, da un punto di vista non orientato all’efficienza-a-tutti-i-costi. Ciò ci condurrà inconsapevolmente alla scoperta di panorami nuovi, di nuovi scorci di realtà, di alterità inaspettate. È l’altro aspetto del mondo accademico che *PageRank* non può prevedere, proprio perché nell’imprevedibilità consiste la sua forza: lasciarsi stupire e scoprirsi ignoranti, e lasciarsi quindi guidare verso percorsi inattesi e certamente arricchenti.

Cambiare o specificare la ricerca, arricchire il proprio punto di vista potrà allora generare un percorso più consono alle esigenze. [...] L’emergenza è un comportamento descrivibile in termini matematici, un’esplosione di complessità inaspettata e imprevedibile, ma soprattutto è un evento in grado di determinare situazioni non descrivibili in maniera esaustiva. Analizzare e percorrere un “universo emergente” come il Web, implica dunque un continuo riposizionamento personale che determina un “mondo chiuso localizzato” di competenze e aspettative, ma anche l’apertura a nuove possibilità di esplorazione (altri mondi sono sempre possibili, fuori dal proprio mondo chiuso) e dunque la consapevolezza di poter raggiungere risultati sempre e solo parziali. (p. 62)

Per l’Autore l’esito culturale di Google sarà - ancora una volta - la “tecnocrazia” (cap. 7), che spoglia lo strumento della caratteristica di *attrezzo* e lo riveste di un’aura di sacralità e infallibilità. L’atteggiamento dell’utente medio, il quale accetta passivamente che l’eccitante processo di ricerca venga fatto da altri e non si riserva il piacere/dovere di cercare da solo, imparando ad affinare i propri strumenti («Solo una scelta di autoformazione può costituire una via di fuga percorribile per sottrarsi al dominio tecnocratico.», p. 76), sarà di acriticità, di attesa che una risposta - inevitabilmente quella *giusta* - arrivi dalla Pizia della modernità. Fin qui l’analisi di Ippolita è politicamente appropriata. Ma il pregio del volume va oltre la mera spiegazione politica. Infatti l’Autore è un collettivo di agitatori sociali (così si definiscono nel sito dell’editore Feltrinelli), ma anche di *hacker* e appassionati di informatica, oltre che disincantati utenti della Rete. Sistematicamente quindi, ma con parole semplici, concetti chiari e spiegazioni esaustive, seppur non pesanti concettualmente, riescono a dare una panoramica su come sia tecnicamente possibile che Google si riveli essere non il Gigante Buono su cui ha puntato il *marketing*, bensì un accentramento di potere preoccupante. Entrano quindi in gioco le modalità di sviluppo del *software* di Google, che usa e appoggia le comunità *Open Source* ma chiude sistematicamente tutto il proprio *software* come proprietario. Si spiega la lotta senza esclusione di colpi tra Microsoft e Google sull’imposizione di *standard*; si analizza approfonditamente il ruolo dei *cookies*, che Google invia alle macchine degli utenti, spesso a loro insaputa, per poterne profilare i gusti, le esigenze, le ricerche, le preferenze. Lungo tutto il (breve) libro l’esposizione politica procede di pari passo con quella tecnica, in modo da non lasciare nessun dubbio che quanto si sta dicendo è uno dei panorami possibili, ossia una delle tante narrazioni che si possono fare analizzando l’attività di Google, e lasciando da parte per un istante quanto il gigante di Mountain View dice di se stesso. Dal metodo di reclutamento della pubblicità alla creazione della “Rete di Google”, dal *PageRank* alla “personificazione di massa”; dalla mossa politico-strategica di includere Vinton Cerf, inventore del TCP/IP e che sedeva nel C.d.A. dell’ICANN (l’organismo che supervisiona il sistema dei nomi e dei domini Internet) quale vicepresidente, alla proposizione di strumenti *on-line* per gli uffici, nel tentativo di sottrarre al mercato degli applicativi da scrivania (leggi Microsoft Office) una grossa fetta di mercato e di spostarlo in Rete, esattamente negli incontrollabili *server* Google.

Ma a questa storia l’Autore non si (op)pone con la cieca fiducia in un mondo perfetto. Ippolita è ben consapevole che Google è solo uno dei soggetti in gioco, e che se non ci fosse Google qualcun

altro (Microsoft, il governo U.S.A., Echelon, etc.) porrebbe le medesime problematiche. Non è questione di soggetti, ma di atteggiamenti. L'ultimo capitolo quindi suggerisce metodi di "autodifesa digitale" e di "autoformazione permanente". Dalla crittografia, al *peer-to-peer* (p2p); dall'uso di *blog aggregator*, alla collaborazione costante presso media indipendenti; dalla condivisione di file, alla contribuzione a *trusting networks*. Tutti modi concreti e non fondamentalisti di opporsi a un nuovo potere di controllo pervasivo e occulto; tutti modi efficaci e critici di tenersi svegli in un panorama di intorpidimento mediatico indotto.

Il testo recensito

Ippolita. *Luci e ombre di Google. Futuro e passato dell'industria dei metadati*. Feltrinelli. Milano. 2007. La versione recensita si basa sul testo disponibile on-line all'indirizzo del gruppo Ippolita. Per ammissione dell'autore il testo on-line non corrisponde esattamente alla versione cartacea.